2012 IEEE International Conference on Multimedia and Expo Workshops

VIRTUAL INTERACTIONS: CAN EEG HELP MAKE THE DIFFERENCE WITH REAL INTERACTION?

Jan Rzepecki, Jonathan Delcourt, Matthieu Perreira Da Silva, Patrick Le Callet

IRCCyN - University of Nantes
La Chantrerie - rue Christian Pauc - BP 50609
44306 Nantes Cedex 03
France

ABSTRACT

Science and technology progress fast, but mouse and keyboard are still used to control multimedia devices. One of the limiting factors of gesture based HCIs adoption is the detection of the user's intention to interact. This study tries to make a step in that direction with use of consumer EEG sensor headset. EEG headset records in real-time data that can help to identify intention of the user based on his emotional state. For each subject EEG responses for different stimuli are recorded. Acquiring these data allows to determine the potential of EEG based intention detection. The findings are promising and with proper implementation should allow to building a new type of HCI devices.

Index Terms— Brain Computer Interface, Electroencephalography, Human Computer Interaction.

I. INTRODUCTION

Providing natural or at least easy and convenient Human Computer Interaction (HCI) or Man-Machine-Interaction (MMI) is a hot feature of many emerging multimedia applications. In the field of computer games, all major brands have developed new interaction devices for their respective consoles: Nintendo Wii-mote, Sony Playstation Move, or Microsoft Kinect. These examples provide good gesture based interaction. Recent advances in pattern matching, machine learning and computer vision allow to extend interaction facilities. Face recognition [1] or speech-recognition [2] can be used to personalize user experience based on its identity, face detection [3], gesture [4] or gaze-tracking can be used as input device for "mouse" pointer replacement, etc. They can even be combined in order to perform multimodal HCI [5].

Despite these scientific and technological progresses, we still generally use old and basic interaction devices and methods in order to communicate with computers and multimedia devices: mouse, keyboards and other simple key based remotes which date back to the early 50's are still standard devices. Even modern smartphones and tablets use the touchscreen paradigm which was developed in the end of the 60's [6]. This could be explained by the fact that users do not adopt new HCI's based on their technological performance, but mostly on their ergonomics, naturalness and ease of use [7]. As a matter of fact, most of efficient HCI are limited to direct contact interface. A recurrent issue is related to the difficulty to identify gestures, vocal inputs and others which are not intended to trigger any interaction with the machine (talking to someone else, taking a drink, moving, etc.). Consequently, one limiting factor of current HCIs is the detection of the intention of the user with respect to the multimedia device [8]. Solving this problem would lead to devices that "know" when you are addressing an order to them and thus would allow much more natural interactions (no/less false alarms on gestures/words not directed to the machine). This is somehow related to attentive user interfaces, which try to adapt the quantity and the way information is provided to users in order to keep their cognitive load at a reasonable value [9].

This paper is focused on the automatic detection of user intention and improvement of comfort of interaction. In particular, we try to determine if there is a perceptible difference on "meta-electroencephalograph" signals between two interaction scenarios: actions in the real world vs actions in relation with an environment displayed on a screen. Another question is related to the manipulation of real or virtual objects. If this distinction is possible, it is the first contribution for more advanced multimodal user intention detection, and so more natural human computer interaction without touch contact. The intention detection that is subject of the study can be linked to works on affective computing because in these two cases physiological signals are used. On the other hand this work focus on intention and affective computing focus on affect[1][10].

In section 2, we first describe experimental protocol that have been adopted to perform user interaction experiments using EEG sensors. As different interaction conditions are considered, the rest of the paper is dedicated to data processing to seek for any automatic classification of these

[1] an expressed or observed emotional response

978-0-7695-4729-9/12 $26.00 © 2012 IEEE
DOI 10.1109/ICMEW.2012.33

conditions based on EEG recordings. Data preprocessing are presented in section 3 prior to describe, in section 4, classification process and performance analysis. Finally conclusions are drawn in last section.

II. EXPERIMENTS AND PROTOCOL

In order to investigate the possibility to distinguish between human interaction with computer (through information displayed on a screen) or interaction with real world, this paper focus on brain activity instead of analyzing all movements, voice commands or other possible interactions with computer. Each interaction takes its origin in brain, it is consequently interesting to check if an EEG headset can lead to access interaction intentions from brain activity (to certain extent, as it will be explained later). Towards this goal, we have designed an experiment based on recording brain signals of test participants via EEG headset. Different user exercises, all based on moving hand across a table following a predesignated rectangular path, have been considered. There are, in total, four type of exercise (with or without object and with or without feedback).

Experiments are based on recording brain signals of test subjects via EEG headset during different activities based on moving hand across a table, following a predesignated rectangular path. There are four variations of the activities : with or without object and with or without feedback.

II-A. Subjects

Experiments have been performed on a population of 12 persons, for whom average age is 26.3 years with a standard deviation of 2.3 years. The population is composed of 10 males and 2 females coming from 6 different countries. All participants are daily computer users.

II-B. Hardware and software

The experiments have been conducted on an Intel Xeon machine with a nVidia Quadro 4000 graphics card, using Microsoft Kinect camera for hand tracking and Emotiv EPOC headset for EEG recording and processing. The testing room where tests have been performed conforms to ITU-R BT500-11 recommendation (fig.1).

Software used consists of an application developed with Unity 3D game engine for visual feedback (fig. 2). A Unity 3D library was also used to record EEG data. Additionally, the Emotiv SDK Control Panel was used to monitor EEG preprocessed data. Finally, hand tracking was made with use of OpenNI framework which provides enough facilities to automatically detect gestures and track hands[2]. Tracking was made via detection of temporal and spatial continuity of the tracked object. The method used is based on depth map camera[3] and is used only for visual feedback (no hand positioning measure are performed).

Instead of recording raw brain signals, preprocessed emotional state features provided by EPOC SDK: Excitement, Engagement/Boredom, Meditation and Frustration have been used. According to the manufacturer these features were extracted from EEG signals as universal for different subjects on a wide range of stimuli, mainly in the game context. These signals are defined as follows:

- **Excitement** describes anticipation for stimulating content, game reward, unexpected event, story twist. It is registered in terms of short-term and long-term excitement.
- **Engagement/Boredom** is linked to how subjects are engaged in current activity. Data was collected during intensive gameplay that was challenging but not overwhelming. Conversely Boredom was registered when little happened in the game.
- **Frustration** represents the level of stress connected with resistance to the fulfillment of certain tasks. Experiments tested response on problems during gaming sessions such as: problems with connectivity and peaks of latency, problems with shot accuracy in first person shooter games.
- **Meditation** is related to relaxation of the subjects. Reference features extraction was made with wide open eye subjects that were trained in meditation technique. Reference data was gathered in sessions of relaxed conversation altering with documentation reading,

II-C. Apparatus and task

Procedure was explained to each subject before starting each test. As it was mentioned before each one had to perform four different tasks connected with different stimuli:

- with object and with feedback (WOWF);
- with no object and with feedback (NOWF);
- with object and with no feedback (WONF);
- with no object and with no feedback (NONF).

Lets us define a visual feedback task (WF) as a test where the subject is kept focused on the screen, on which he sees a virtual object following his hand movements. Depending on which state of experiment takes place, virtual object is a shape of a ball - for movements not recorded, simulating hand position - or a cylinder - for recorded movements, simulating moving object. The use of these two objects try to help a test subject to differentiate when he virtually drags an object or not.

No visual feedback task (NF) is a test where the subject is kept focused on his hand and natural environment (real world). During that test the feedback screen is switched off.

[2]in this project OpenNI is installed with the component for Unity via ZigFu developer bundle : http://zigfu.com/legacy.html

[3]Kinect device can provide three signals: depth map video, RGB video and multichannel audio signal.

The other part of the experiment is when a test subject holds an object or not. Holding object task (WO) is defined as a task where the subject grabs a cylindrical object in his hand (in our test the object is a small plastic bottle). In contrast, empty hand task (NO) is when the subject holds nothing - but makes an action of holding something (imaginary cylinder for example).

The activity performed in all tasks is the same and is based on hand movement between four markers put on a table in a rectangular shape. Movements are ordered in clockwise direction, slightly raising hand between markers and laying down on them (a good precision is not asked to test subjects). The activity is repeated continuously for 30 seconds. In cases where subjects have no visual feedback, end of exercise is signalized by supervisor based on timer included in the application.

Each subject performed the experiment as follows:

- setting up EEG headset;
- one round of the four different tasks is performed by the test subject to see if he has understood the instructions (and is helped if he has not);
- finally the tasks recording starts

Datasets recordings were made repeating 4 times the same task (giving 16 datasets for the 4 tasks). Number of acquisitions for the same user was selected to 4 as optimal compromise between largest possible amount of data and mental fatigue of subjects. Due to technical issues described later on, the number of datasets for some users had to be reduced to 12 or 8. Consequently users were divided into 3 groups based upon number of datasets available, ranging from 2 to 4 datasets per task. The final distribution of the number of recorded datasets is as follows:

- 4 users with 16 datasets (4 times the 4 tasks),
- 6 users with 12 datasets (3 times the 4 tasks),
- 2 users with 8 datasets (2 times the 4 tasks).

For data analysis, we added users with a greater number of datasets to those with the smaller number of datasets by removing one dataset for each experiment. It resulted in 3 different groups of datasets:

- G1 (group 1): 4 users with 16 datasets,
- G2: 10 users with 12 datasets
- G3: 12 users with 8 datasets.

III. DATA PREPROCESSING

Before EEG data can be analyzed, some pre-processing is necessary. Indeed, due to technical issues some of recorded data had to be cleaned or, depending on test, datasets were normalized.

III-A. Technical issues

It occurred that during the acquisition process the EEG headset gave invalid or no data. It appears that the problems was coming from two main causes:

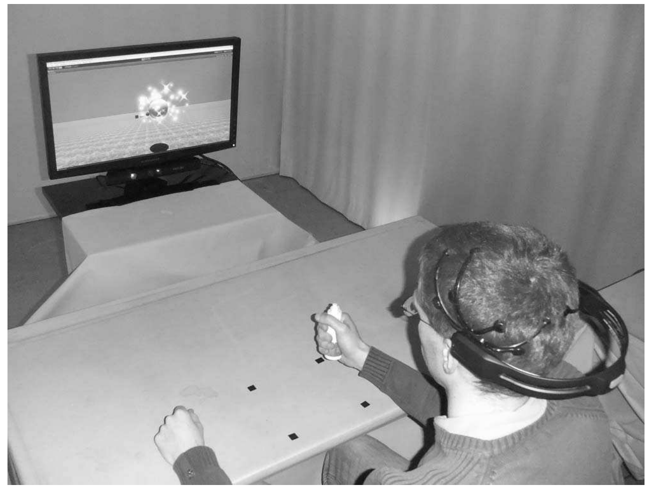

Fig. 1. Testing room and experiments setup.

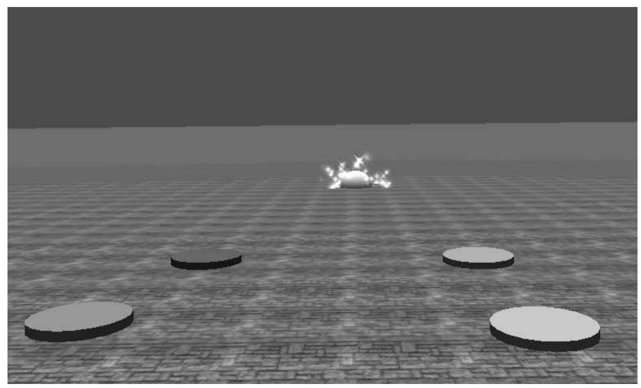

Fig. 2. Preview of the visual feedback window presented to the subject during the experiments.

- **wireless connection** was not stable causing data loss during transmission. In that case the last received value of the signal was repeated (and information indicating that this acquisition is invalid is also stored).
- **high sensitivity of sensor** used to capture electric activity of the brain can also detect electric noise, especially muscles electric activity. So when the user speaks or frowns, sensors could record wrong data.

Another observed error was that headset state sometimes became flickering. In that case the same data is recorded twice. All datasets were checked and cleaned from invalid data using automated scripts.

III-B. Normalization

Our experiments are based on pre-processed data registered from EPOC device SDK. The EEG device can give inter-user normalized data, but normalization procedure is not known and (according to the manufacturer documentation) can only be effective after 30 minutes of brain signal recording.

In the first part of our study (intra-user) we do not need to normalize data, since analysis is done for each user independently. However, for the second part of the study (inter-user comparison) we need to make a proper normalization to ensure that all data are comparable. Mean values and standard deviation of the same attributes differ between different subjects. However these differences are constant between experiments, so data are normalized by transforming the mean value of each user to *zero* and rescaling by adjusting standard deviation to *one* (we make the hypothesis that the data have a normal distribution) . This operation is defined by equation 1

$$x_k^N = \frac{x_k - \bar{x}}{\sigma_x} \tag{1}$$

where x_k is a series of a single attribute acquired by EEG headset, $\bar{x}$ and σ_x are mean and standard deviation values of the attribute for all data gathered for that user.

IV. DATA ANALYSIS

This section is composed of 3 different analysis. Firstly, users are taken into account independently. EEG signals are classified for each user with no regard to other users' datasets. A sampling length study is also conducted to find out if the sample size influences the classification accuracy. Secondly, all user datasets are analyzed together aiming to verify the possibility of building a universal model. Finally, we study the importance of the five EEG signals and their impact on the classification accuracy. Orange Canvas[4] and Numpy[5], respectively, a Python based data visualization and analysis tool for data mining and a Python package for scientific computing, are used in this section to calculate and visualize different classifiers on our datasets.

In the rest of this paper 8 classes will be used, divided into 3 groups resuming the definitions of section II-C:

- a group of 4 classes:
  NONF, NOWF, WONF and WOWF (type OF).
- two groups of 2 classes:
  - NO and WO (type O).
  - NF and WF (type F).

These groups were constituted to serve different purpose. First of all we wanted to be able to classify all type of activities (OF). Later in search for increase in classification accuracy groups O and F were formed, grouping together classes describing opposite activities (with-object, no-object)

IV-A. Intra-user analysis

The following analysis is focused on single user datasets, aiming to classify subject activity in 4 classes – type OF (see section II-C).

[4]http://orange.biolab.si/

[5]http://numpy.scipy.org/

Table I. Classification accuracy using the SVM classifier (1 sample, different types of classification)

		Type		
		O	**F**	**OF**
G1	$\overline{CA}$	96.7%	98.1%	96.6%
	σ	1.73%	1.26%	1.57%
G2	$\overline{CA}$	98.3%	99.1%	98.2%
	σ	0.98%	1.25%	0.99%
G3	$\overline{CA}$	98.9%	99.5%	99.1%
	σ	0.83%	0.56%	0.74%

Classification is made with 4 different methods which are SVM[6], k-NN[7], Naive Bayes and Tree Classifier.. Parameters of these classifiers are presented below.

SVM	is implemented in libsvm library using C-SVM type with RBF
k-NN	is using 5 neighbors, euclidean metric and normalized continuous attributes.
Naive Bayes	is using relative frequency method for probability estimation
Tree Classifier	with attribute selection based upon Information Gain.

For other parameters, classifiers use default values supplied by Orange Canvas. Each classifier is tested with random sampling for training and testing data (repeated 10 times with 70:30% relative size ratio). Performance of classification is measured by means of classification accuracy (CA).

Data is analyzed dividing it into smaller chunks of (1, 5, 10, 15) samples. For *avg* or *mean* analysis, chunks are later averaged over values of the same attribute. For raw analysis, attributes from different samples form an input vector of $k*n$ size (k being the number of samples and $n = 5$ the number of attributes).

Results show the same tendency for all users under all groups and for each classifier. Results in function of different sizes of chunks, and *mean*/*raw* types of features are depicted in Fig. 3 and 4 (only results of SVM classifier for G1 and G2 are shown for illustration). Best classification results are obtained for 1 sample chunks, CA decreases with increasing number of samples. Reducing the number of attributes by averaging helps slightly, however the accuracy is still not as good as for lower sample count. The SVM classifier gives better results than other classifiers (naive Bayes being the worst), but cost little more time for training. In some applications, 4 classes are not necessary, so more general classes can be used , table I presents results of SVM classification for different groups of classes (O, F, OF). We can observe that F type classification gives better results than two others for all analyzed groups.

The results obtained can be explained by the fact that averaging masks signal fluctuations, which appear in raw

[6]Support Vector Machine

[7]k-Nearest Neighbors

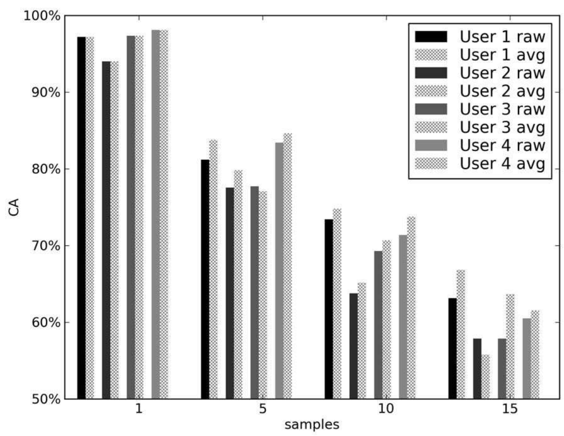

Fig. 3. Classification accuracy for the 4 users of group G1, using SVM classifier

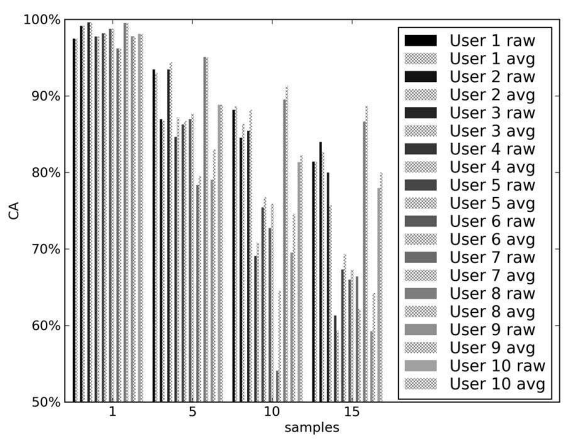

Fig. 4. Classification accuracy for the 10 users of group G2, using SVM classifier

samples - and so slightly improve classification results. Because classification results are better for 1 sample, we can make near real-time classification (limited to sampling acquisition speed and SVM classification one). Slightly better results, for F type of classification can be explained by lower number of classes. However, the same case is for O type of classification, and it notes lower results. Therefore, it is more probable that visual feedback stimuli have greater impact on brain data registered by EEG headset.

IV-B. Inter-user analysis

The aim of this analysis is to assess the feasibility of a universal (inter-user) classifier, which posses pre-trained classes that are efficient regardless of users.

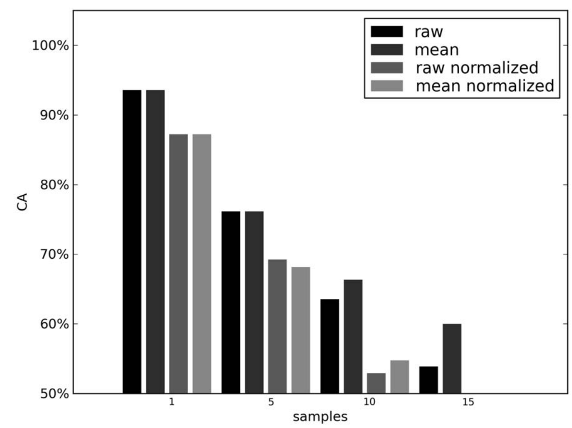

Fig. 5. Classification results for a user independent model on group G1

Figure 5 presents results showing the relation between classification accuracy and two factors: the number of chunks and the type of processing of data for G1. Processing consists of the two processes described before: reduction by averaging and normalizing datasets with users' statistical parameters (see sec. III-B). Results are still not as good as in intra-user analysis, but we can observe the same tendencies. Another conclusion that can be drawn from these results is that the normalization method used here is not well suited to this kind of data and classification.

IV-C. Attributes importance analysis

The aim of this analysis is to quantify the importance of the 5 different EEG signals on classification.

Features used are limited to 1 sample raw data chunks (with no averaging) and SVM classifier.

Analysis is based upon evaluating all possible combinations of attributes in search for CA parameter. For every attribute (Long Term Excitement, Short Term Excitement, Frustration, Boredom, Meditation) all combinations (along each group of users) including this attribute were listed. CA scores were paired so CA score for subset including certain attribute and CA score for the same subset but with this attribute removed were taken together. Each CA score formed as many pairs as it was necessary. Differences of scores in pairs were taken and averaged between different pairs of same type of attribute forming attribute's averaged loss of classification $CAL(s_i)$.

$$S = \{lte,\, ste,\, fru,\, bor,\, med\} \;\; T = 2^S \backslash \{\varnothing\} \qquad (2)$$

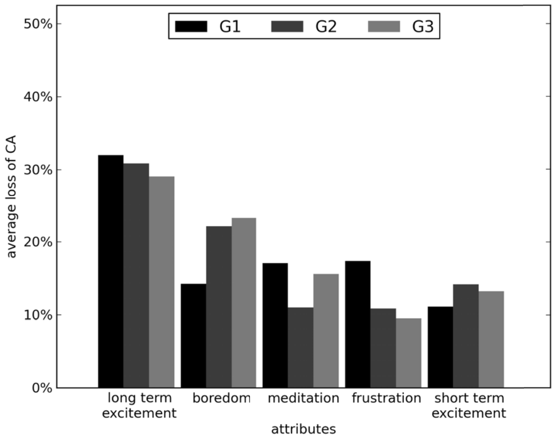

Fig. 6. Impact of each emotional state feature on classification accuracy. Average loss of classification accuracy on attribute removal.

$$CAL(s_i) = \frac{\sum_{j=0}^{|T|} \begin{cases} CA(t_j) - CA(t_j \backslash \{s_i\}) & if\, s_i \in t_j \\ 0 & else \end{cases}}{\sum_{j=0}^{|T|} \begin{cases} 1 & if\, s_i \in t_j \\ 0 & else \end{cases}} \tag{3}$$

The results obtained are presented in fig.6. Despite the fact that values varies slightly between groups, it is noteworthy that Long Term Excitement attribute and later Boredom attribute have higher importance than other attributes, it means that by their removal classification accuracy drops on average by the highest percentage.

V. CONCLUSION

EEG based classification of user interaction is the first step for better HCI devices, as it opens more natural ways of controlling machines. The results obtained from the experiment are promising, especially for intra-user analysis we found that classification accuracy is really high (CA 96%). It can be further improved by reducing the number of classes. Still it is surprising that even considering high level features used in experiment, results are so good. Considering 1 sample classification requirement (0.5 s), the process can be nearly real-time. This can possibly lead to HCI device that detects intentions and instantaneously knows when user wants to interact with it. To sum up, this work opens the door for further investigation.

VI. REFERENCES

[1] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *Acm Computing Surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003. [Online]. Available: http://dl.acm.org/citation.cfm?id=954342

[2] M. A. Anusuya and S. K. Katti, "Speech Recognition by Machine : A Review," *Journal of Computer Science and Information Security*, vol. 6, no. 3, pp. 181–205, 2009.

[3] C. Zhang, "A survey of recent advances in face detection," *Microsoft Research, June*, no. June, 2010. [Online]. Available: http://research.microsoft.com/en-us/um/people/chazhang/publications/tr2010-66_chazhang.pdf

[4] S. Mitra, "Gesture recognition: A survey," *Systems, Man, and Cybernetics, Part C:*, vol. 37, no. 3, pp. 311–324, 2007. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4154947

[5] A. Jaimes and N. Sebe, "Multimodal human–computer interaction: A survey," *Computer Vision and Image Understanding*, vol. 108, no. 1-2, pp. 116–134, Oct. 2007. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1077314206002335

[6] E. A. Johnson, "Touch display—a novel input/output device for computers," *Electronics Letters*, vol. 1, no. 8, p. 219, 1965. [Online]. Available: http://link.aip.org/link/ELLEAK/v1/i8/p219/s2\&Agg=doi

[7] V. Venkatesh, M. Morris, G. Davis, and F. Davis, "User acceptance of information technology: Toward a unified view," *MIS quarterly*, vol. 27, no. 3, pp. 425–478, 2003. [Online]. Available: http://www.jstor.org/stable/10.2307/30036540

[8] Z. Bien, K. Park, and J. Jung, "Intention reading is essential in human-friendly interfaces for the elderly and the handicapped," *Industrial Electronics, IEEE*, vol. 52, no. 6, pp. 1500–1505, 2005. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1546365

[9] R. Vertegaal, "Attentive user interfaces," *Communications of the ACM*, vol. 46, no. 3, pp. 30–33, 2003. [Online]. Available: http://www.hml.queensu.ca/papers/vertegaalcacm0303.pdf

[10] R. Picard, "Affective computing. media laboratory, perceptual computing tr 321," 1995.